

REMARKS

Claims 1-10 stand in the present application. Reconsideration and favorable action is respectfully requested in view of the following remarks.

In the Office Action, the Examiner has rejected claims 1-10 under 35 U.S.C. § 102(b) as being anticipated by Choi. Applicants respectfully traverse the Examiner's § 102 rejection of the claims.

Applicants' invention involves an information retrieval system that recognizes the semantic similarity of different words to determine whether one document has similar contents to another. For example, with no semantic knowledge or understanding a computer will not match the word "taxi" with the word "cab."

As noted previously, in Applicants' invention it is possible for the system to determine the semantic similarity of words; the text of the document set is stemmed, optionally after the exclusion of the most and least common words from the document set, and the resulting word output is analyzed to determine a number of n-grams. Paragraphs [0055] - [0063] of the present application provide an example of how four sentences may be analyzed to generate a number of 3-grams (that is an n-gram for the case where $n=3$). Conversely, in Choi it is necessary for a human user to submit a set of keywords that have a semantic similarity in order that matches between different documents can be determined, for example to tell the system that "taxi" and "cab" have the same meaning.

The Office Action states that "[a]lthough the applicant's specification discloses the stemming algorithm, it is not mentioned in the claims." To the contrary each of independent claims 1, 3, 8, and 10 recite limitations directed to the stemming algorithm

– which limitations are not found in the cited reference. For example, integers of claim 1 directed to the stemming algorithm include:

- (i) for each word of said plurality of words:
 - (a) identifying, in documents of said set of one or more documents, word sequences comprising the word and a predetermined number of other words;
 - (b) calculating a relative frequency of occurrence for each distinct word sequence among word sequences containing the word; and
 - (c) generating a fuzzy set comprising, for word sequences containing the word, corresponding fuzzy membership values calculated from the relative frequencies determined at step (b) . . .

See, claim 1 (emphasis supplied).

The Office Action states that limitation i(a) can be found at paragraph [0002] of Choi. See, Office Action at page 4.

The present invention relates to a method of order-ranking document clusters using entropy data and Bayesian self-organizing feature maps(SOM), in which an accuracy of information retrieval is improved by adopting Bayesian SOM for performing a real-time document clustering for relevant documents in accordance with a degree of semantic similarity between entropy data extracted using entropy value and user profiles and query words given by a user, wherein the Bayesian SOM is a combination of Bayesian statistical technique and Kohonen network that is a type of an unsupervised learning.

However, the cited portion of Choi does not disclose “for each word of said –plurality of words . . . identifying, in documents . . . word sequences comprising the word and a predetermined number of other words” as required by claim 1. Choi merely discloses the conventional use of query words, and does not even mention identifying word

sequences comprising the word and a predetermined number of other words. The Office Action misapprehends Applicants' invention by asserting:

"Finding word sequences" is anticipated by, "query words given by the user" (§ 0002).

Id. In fact, the above quotation is not a limitation of claim 1 – the actual limitation is "identifying . . . word sequences comprising the word and a predetermined number of other words." Nowhere does Choi teach or suggest identifying word sequences comprising its query word and a predetermined number of other words. Thus, Choi fails to teach or suggest limitation i(a).

The Office Action suggests that limitation i(b) can be found at paragraph [0027] of Choi, stating that portion of Choi discloses "search and frequencies of keywords." *Id.*

To accomplish the above objects of the present invention, there is provided a method of order-ranking document clusters using entropy data and Bayesian SOM, including a first step of recording a query word by a user; a second step of designing a user profile made up of keywords used for the most recent search and frequencies of the keywords, so as to reflect a user's preference; a third step of calculating entropy value between keywords of each web document and the query word and user profile; a fourth step of judging whether data for learning Kohonen neural network which is a type of unsupervised neural network model, is sufficient or not; a fifth step of ensuring the number of documents using a bootstrap algorithm, a type of statistical technique, if it is determined in the fourth step that the data for learning Kohonen neural network is not sufficient; a sixth step of determining prior information to be used as an initial value for each parameter of network through Bayesian learning, and determining an initial connection weight value of Bayesian SOM neural network model where the Kohonen neural network and Bayesian learning are coupled one another; and a seventh step of performing a real-time document clustering for relevant documents using the

entropy value calculated in the third step and Bayesian SOM neural network model.

However, the cited portion of Choi does not disclose the actual limitation of claim 1, i.e., “calculating a relative frequency of occurrence for each distinct word sequence among word sequences containing the word.” Indeed, the cited portion of Choi discloses nothing about frequencies involving each distinct word sequence among word sequences containing the words, as required by the claim, but merely frequencies of the keywords. Thus, Choi fails to teach or suggest limitation i(b).

The previous Office Action suggested that limitation i(c) can be found at paragraph [0143] of Choi.

Clustering method includes k-nearest neighbor method, fuzzy method and the like. However, the present invention adopts a clustering method where documents are clustered by a statistical similarity, i.e., standardized distance between the two documents. In other words, a hierarchical document clustering where document cluster is formed through grouping documents having high statistical similarity, starting from each clusters made up of each documents expressed in terms of statistical similarity.

See, Office Action, dated November 28, 2007, at page 3. However, the cited portion of Choi does not disclose the actual limitation of claim 1, i.e., “generating a fuzzy set comprising, for word sequences containing the word, corresponding fuzzy membership values” Indeed, once again, the cited portion of Choi makes no mention of word sequences at all. Thus, Choi fails to teach or suggest limitation i(c).

As demonstrated above, the claim integers (i) (a) - (c) are not disclosed (or even suggested) by Choi and, thus, claim 1 and its dependent claims patentably define over Choi. Independent claims 3, 6, 8, and 10 contain similar claim integers and, therefore,

these claims and their respective dependent claims also patentably define over Choi.

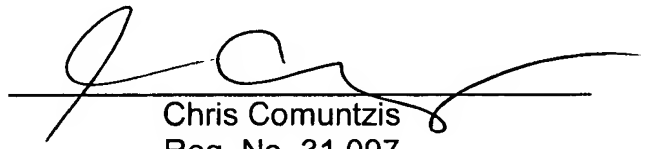
As outlined previously, Choi does not teach towards the solution provided by Applicants' invention; indeed Choi discloses the use of a manual method of providing a semantic link and thus Choi teaches **away** from the present invention.

Therefore, in view of the above remarks, it is respectfully requested that the application be reconsidered and that all of claims 1-10, standing in the application, be allowed and that the case be passed to issue. If there are any other issues remaining which the Examiner believes could be resolved through either a supplemental response amendment after final rejection

Respectfully submitted,

NIXON & VANDERHYE P.C.

By: _____


Chris Comuntzis
Reg. No. 31,097

CC:Imr
901 North Glebe Road, 11th Floor
Arlington, VA 22203-1808
Telephone: (703) 816-4000
Facsimile: (703) 816-4100